Patent Office
Research and Development
Reports . . . . . No. 11

A MANUAL FOR CODING STEROIDS

Prepared by

Julius Frome, Jacob Leibowitz
Patent Research Specialists

Office of Research and Development
Patent Office

November 17, 1958

Robert C. Watson
Commissioner of Patents

Lewis L. Strauss
Secretary of Commerce

For sale by the U. S. Department of Commerce, Washington 25, D. C. Price 25 cents.

Table of Contents

Acknowledgment

Acknowledgment is expressed for the contributions of Mrs. Rowena W. Swanson, patent examiner, Mechanized Examining Division A, in editing and revising the manuscript, and Mr. Walter A. Modance, patent examiner, Mechanized Examining Division A, in making substantive coding suggestions.

A MANUAL FOR CODING STEROIDS

INTRODUCTION

The system devised by the Office of Research and Development (R & D) of the U. S. Patent Office for the mechanized searching of steroid compounds is described in R & D Report No. 7 entitled *A Punched Card System for Searching Steroid Compounds.*[1]

SCOPE OF THE ART IN THE SYSTEM

The system is limited to the steroid art. The patents included in the system are those classified in Class 260, Subclasses 239.5, 239.55, 239.57, 397, 397.1, 397.2, 397.25, 397.3, 397.4, 397.45, 397.47, and 397.5.[2]

Only those steroids disclosed in the patents which meet the definitions of the above Class and Subclasses have been included in the system.[3]

The seco and homo steroids are excluded. Also excluded are steroids classified elsewhere according to the Rule of Superiority in classification.[4]

GENERAL CODING PRINCIPLES

Every compound which is coded must contain the steroid nucleus shown in Figure 1. A fixed numbering system is assigned to this nucleus (see Figure 4), the fixed numbers serving to identify the positional locations of nuclear substitution.

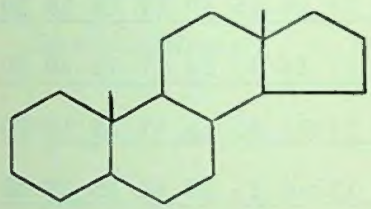


Figure 1

Coding of the patents is done on a one-page form (front and back) on which the terms representing information to be coded with their corresponding codes are assembled. The top side of the form is designated "2A" and is shown in Figure 2. The reverse side is designated "2B" and is shown in Figure 3. The terms are called "descriptors." See pages 6 and 7.

Composite Coding

Composite coding is described in R & D Report No. 7. Briefly, a group of formulas of related compounds are "composited" into a single synthetic formula to reduce the number of total codes per document.

Multiple Coding

The terms which have been selected as descriptors present a certain degree of overlap in concept. Thus a compound or group of compounds may be describable by more than one term. However, because the terms represent various levels of specificity (or genericity), all terms which are applicable to a particular configuration are used.

Relationship of Codes and Punched-Card Format

Each descriptor is defined by a single set of numbers designating a particular column and row of the standard 80-column IBM punched card. Conversely, each column-row location on the IBM card is a fixed allocated space for a particular descriptor. The card is divided into two fields: columns 1 through 59 constituting the field for 2A terms and columns 60 through 67 constituting the field for 2B terms, more fully described below.

NUMBERING SYSTEM FOR THE NUCLEUS

The following fixed numbering system is assigned to the steroid nucleus:

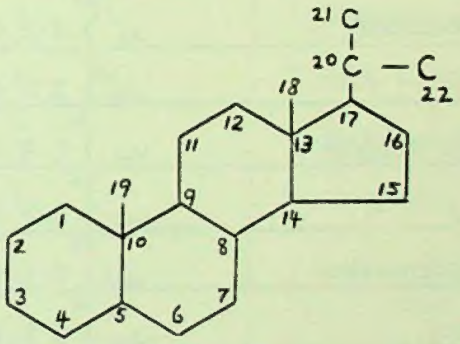


Figure 4

Positional locations 18 through 22 are variables in the sense that they may not be present in a particular compound. The locations exist only when they are occupied by carbon atoms.

ORGANIZATIONAL ARRANGEMENT OF TERMS

2A Terms—General Description

The 2A terms relate to chemical configurations and substituents on the steroid nucleus (Figure 4). The codes for these terms have been devised to show the positional locations of the descriptors on the nucleus. Since there are 22 possible positional locations, two columns on the IBM punched card have been allocated to each 2A descriptor. Thus, the first two columns are assigned to each 2A descriptor. Thus, the first two columns are assigned

Patent No.....Examiner.....Punched.....																								
	O																						11	12
═══	22	1	2	3	4	5	6	7	8	9	2	10	11	12	13	14	15	16	17	18	19	20	21	
∝ or allo	22	3	1	2	3	4	5	6	7	8	9	4	10	11	12	13	14	15	16	17	18	19	20	21
-C≡C.	22	5	1	2	3	4	5	6	7	8	9	6	10	11	12	13	14	15	16	17	18	19	20	21
CH_3	22	7	1	2	3	4	5	6	7	8	9	8	10	11	12	13	14	15	16	17	18	19	20	21
CN	22	9	1	2	3	4	5	6	7	8	9	10	10	11	12	13	14	15	16	17	18	19	20	21
COOH or COOR	22	11	1	2	3	4	5	6	7	8	9	12	10	11	12	13	14	15	16	17	18	19	20	21
-C- sub	22	13	1	2	3	4	5	6	7	8	9	14	10	11	12	13	14	15	16	17	18	19	20	21
-H	22	15	1	2	3	4	5	6	7	8	9	16	10	11	12	13	14	15	16	17	18	19	20	21
NH_2 or N<	22	17	1	2	3	4	5	6	7	8	9	18	10	11	12	13	14	15	16	17	18	19	20	21
OH	22	19	1	2	3	4	5	6	7	8	9	20	10	11	12	13	14	15	16	17	18	19	20	21
=O	22	21	1	2	3	4	5	6	7	8	9	22	10	11	12	13	14	15	16	17	18	19	20	21
-Se-	22	23	1	2	3	4	5	6	7	8	9	24	10	11	12	13	14	15	16	17	18	19	20	21
-S-R	22	25	1	2	3	4	5	6	7	8	9	26	10	11	12	13	14	15	16	17	18	19	20	21
Hal	22	27	1	2	3	4	5	6	7	8	9	28	10	11	12	13	14	15	16	17	18	19	20	21
Hydrocarbon	22	29	1	2	3	4	5	6	7	8	9	30	10	11	12	13	14	15	16	17	18	19	20	21
Ketal	22	31	1	2	3	4	5	6	7	8	9	32	10	11	12	13	14	15	16	17	18	19	20	21
Ketone reagent	22	33	1	2	3	4	5	6	7	8	9	34	10	11	12	13	14	15	16	17	18	19	20	21
Epoxy	22	35	1	2	3	4	5	6	7	8	9	36	10	11	12	13	14	15	16	17	18	19	20	21
-O hydrocarbon	22	37	1	2	3	4	5	6	7	8	9	38	10	11	12	13	14	15	16	17	18	19	20	21
-O Acyl	22	39	1	2	3	4	5	6	7	8	9	40	10	11	12	13	14	15	16	17	18	19	20	21
-O-hetero	22	41	1	2	3	4	5	6	7	8	9	42	10	11	12	13	14	15	16	17	18	19	20	21
-N-hetero	22	43	1	2	3	4	5	6	7	8	9	44	10	11	12	13	14	15	16	17	18	19	20	21
S-hetero	22	45	1	2	3	4	5	6	7	8	9	46	10	11	12	13	14	15	16	17	18	19	20	21
Miscellaneous	22	47	1	2	3	4	5	6	7	8	9	48	10	11	12	13	14	15	16	17	18	19	20	21

Figure 2

60		61		62		63	
0	O-Acyl	0	O-Hetero	0	N-Hetero	0	S-Hetero
1	Carboxylic	1	Morpholine	1	Morpholine	1	Thiophene
2	Poly	2	Furan	2	Piperidine	2	Thiazole
3	Unsat	3	Lactone	3	Pyridine	3	
4	Aromatic	4	Spirostane	4	Pyrimidine	4	
5	Aliphatic	5	Sub in O spiro ring	5	Pyrrole	5	
6	Substi.	6	Psuedosapo.	6	Thiazole	6	
7	St. chain	7		7		7	
8	Cycloalkyl	8		8		8	
9	Branched	9	Misc.	9	Misc.	9	Misc.
11	Heterocyclic	11		11		11	
12	Inorganic except hal						

64		65		66		67	
0	Bile cpds.	0	Sterols	0	Hal.	0	
1	Acids	1	Ergosterol	1	Fl.	1	Androstane
2	Cholanic	2	Cholesterol	2	Br.	2	Add. compd.
3	Norcholanic	3	Vitamin D_3	3	I	3	Maleic adduct
4	Bisnorcholanic	4		4	Cl Double bond	4	Pregnane
5		5		5	5 (6)	5	21 Unsubsti-tuted
6		6		6	5 (10)	6	21 Diazo
7		7		7	8 (9)	7	
8		8		8	8 (14)	8	
9	Misc.	9	Misc.	9	1 (2)	9	Misc.
11		11		11	1 (10)	11	
12		12		12		12	

Figure 3

to "═", the next two columns to "∝ or allo", etc. (see Figure 2). Positional location 1 on the nucleus is coded in row 1 on the IBM punched card, location 2 in row 2, etc.

Figure 5 demonstrates the use of the columns and rows of the IBM card for the 2A term "═". The code for "═" is tabulated for each of the 22 positional locations on the nucleus. Figure 5 also shows the correspondence of the rows of the IBM card with the positional locations.

Positional Location	CODE	
	IBM Card Column	IBM Card Row
1	1	1
2	1	2
3	1	3
4	1	4
5	1	5
6	1	6
7	1	7
8	1	8
9	1	9
10	2	0
11	2	1
12	2	2
13	2	3
14	2	4
15	2	5
16	2	6
17	2	7
18	2	8
19	2	9
20	2	11
21	2	12
22	1	0

Figure 5

Note: The first column of a pair of columns for a 2A term is used for positional locations 1 to 9 and 22; the second column is used for locations 10 to 21. Rows 1 to 9 of the first column correspond to positional locations 0 to 9, respectively; row 0 corresponds to location 22. Rows 0 to 9 of the second column correspond to positional locations 10 to 19, respectively, row 11 to location 20, and row 12 to location 21.

2B Terms—General Description

The 2B terms are, with a few exceptions, broader and more generalized in scope. The terms incorporate generic and subgeneric concepts related to the compound as a whole, e. g., "N-hetero," and, as types of "N-hetero" compounds, "morpholine," "piperidine," etc. The terms are not associated with a positional location. The column and row numbers allocated to these terms are shown in Figure 3, the column numbers being given at the top of each group and the row numbers being listed thereunder. The 2B terms corresponding to the row numbers appear to the right of the numbers.

Pregnanes and Androstanes

When the steroid nucleus has a chain of 2 carbons in single or double bond linkage as in Figure 6,

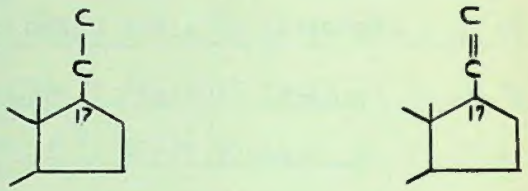


Figure 6

the compound containing this structure is coded as a pregnane. If the 2 carbon chain is in triple bond linkage as in Figure 7,

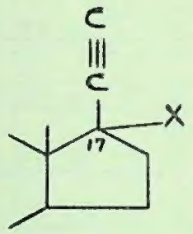


Figure 7

the compound containing the structure is coded as an androstane, unless X represents the linkages shown in Figure 6.

More generally, androstanes are structures which have a non-hydrocarbon substituent, a methyl group, a non-hydrocarbon substituted methyl group, or an ethinyl linkage in the 17 position.

Structures containing groups characteristic of both the pregnanes and androstanes as defined above are coded as pregnanes only, i.e., the androstane structure is subordinated to the pregnane. An example of such a structure is given in Figure 8.

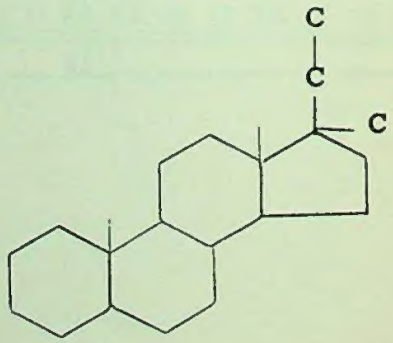


Figure 8

DEFINITIONS OF DESCRIPTORS

The definitions which follow have not necessarily been assigned from the point of view of strict adherence to chemical principles but from the point of view of permitting retrieval of desired classes of steroid compounds.

Definitions—2A Terms

═

This symbol designates a double bond in the nucleus. In coding the positional location of a double bond, the lower number of the pair of position number is indicated. Thus, for $\Delta^{4,5}$ position 4 is coded. Where any of positions 1, 5 and 8 are involved, 2B terms are added to distinguish between positions 1-2, 1-10, 5-6, 5-10, 8-9, 8-14.

∝ or allo

This term represents isomeric forms of a structure. In the case of the 3, 5 or i position in Ring A, as in i-cholesterol, the codes are assigned as though two positional locations were occupied.

Example of coding:

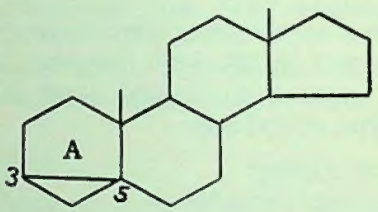


Figure 9

Codes: Column 3 Row 3
Column 3 Row 5
} (Allo)

-C═C-

This is the ethinyl linkage.

Example of coding:
(a 17-ethinyl androstane)

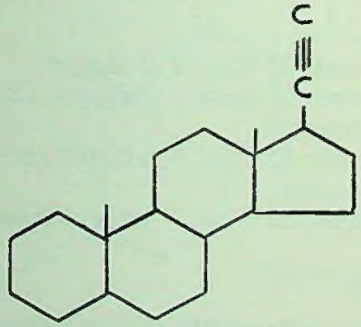


Figure 10

Codes: Column 6 Row 7 (ethinyl)
Column 67 Row 1 (androstane)

CH_3

This represents a methyl group. Methyl groups which appear in positional locations 18, 19, 20, and 21, when present, are considered as integral portions of the nucleus and are not coded as methyl substituents. When a methyl group is coded, the 2A term hydrocarbon is also coded.

Example of coding:
(a 17-methyl pregnane)

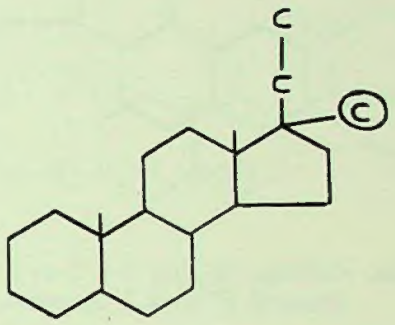


Figure 11

In Figure 11, the methyl group in position 17 only (encircled) is coded as follows:

Codes: Column 8 Row 7 (methyl group)
Column 30 Row 7 (hydrocarbon)
Column 67 Row 4 (pregnane)

CN

CN represents the nitrile group and may be coded with respect to any of the 22 positions.

COOH or COOR

COOH and COOR refer to carboxylic acids and their salts or esters, respectively, where the carboxyl group is directly attached to the nucleus, and may be coded in any of the 22 positions.

Example of coding:
(a methyl ester of 17-carboxy androstane)

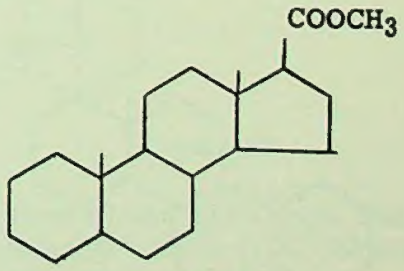


Figure 12

Codes: Column 12 Row 7 (COOH or COOR)
Column 67 Row 1 (androstane)

-C-sub.

This symbol represents a non-hydrocarbon substituent linked through a carbon to the steroid nucleus and not specifically provided for on the list.

Example of coding:
(3-chloro methyl androstane)

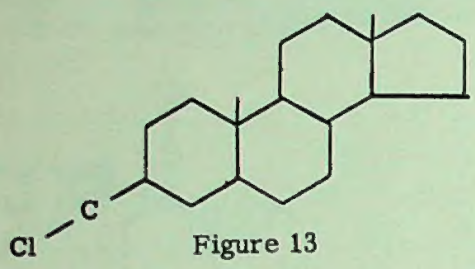


Figure 13

Codes: Column 13 Row 3 (-C-sub.)
Column 67 Row 1 (androstane)

-H

This symbol is applied only when the hydrogen atom is present in the 10 or 13 position or both to indicate hydrogen in place of C_{18} or C_{19} or both.

NH_2 or N<

This symbol represents an amino or substituted amino group. The group is not applicable when the nitrogen atom is part of a nitro group or a heterocyclic ring.

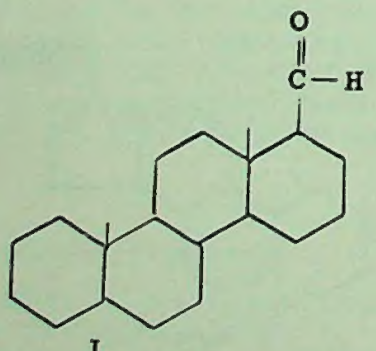
OH

This represents the hydroxy group.

=O

The =O represents the ketone or aldehyde group.

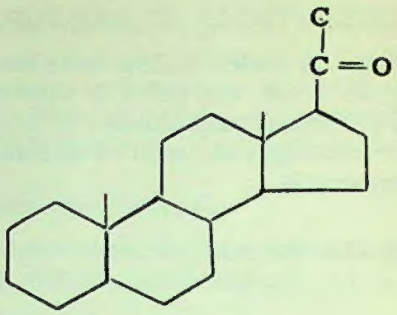
Examples of coding:



I

Codes: Column 22 Row 11 (aldehyde)
Column 67 Row 1 (androstane)

Figure 14



II

Codes: Column 22 Row 11 (ketone)
Column 67 Row 4 (pregnane)

Figure 14—Con.

-Se-

This symbol represents selenium or substituted selenium where the selenium is directly attached to the steroid nucleus.

-S-R

This symbol refers to sulfur or substituted sulfur where the sulfur is directly attached to the steroid nucleus.

Hal

This symbol represents the halogen group. Halogens are coded in 2A with respect to positions. Column 60, rows 0 to 4, are used to further describe the type of halogen.

Example of coding:
(a 4-chloro androstane)

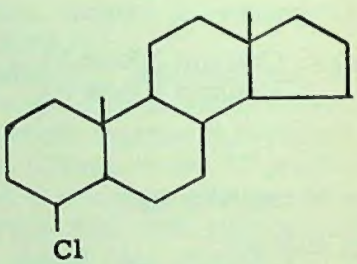


Figure 15

Codes: Column 27 Row 4 (halogen)
Column 66 Row 0 (halogen containing, broadly)
Column 66 Row 4 (chloro containing, broadly)
Column 67 Row 1 (androstane)

Hydrocarbon

This term represents hydrocarbon in any of the positions 1 to 22. Where the hydrocarbon group is a methyl 18 or a methyl 19, it is not coded. Also, in pregnanes, the chain of 2 carbons attached to position 17 is not coded as a hydrocarbon. In

cholesterols and other sterols, the hydrocarbon chain in the 17, 20 and 22 positions is not coded as hydrocarbon.

Ketal

This designation refers to the reaction product resulting from the reaction of a keto or aldehyde group with an alcohol to give a ketal or cyclic ketal. The term "ketal" includes also the term "acetal."

Example of coding:
(a 3, 20-diethylene ketal pregnane)

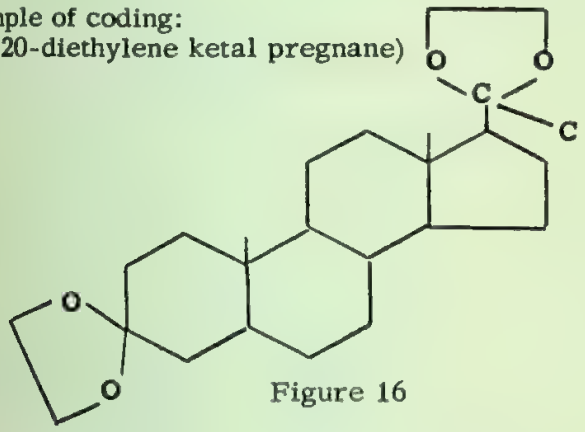


Figure 16

Codes: Column 31 Row 3 (ketal, location 3)
Column 32 Row 11 (ketal, location 20)
Column 61 Row 0 (O-hetero compound)

The last code, 61-0, is included since the cyclic ketal is an oxygen-containing heterocycle.

Ketone reagent

This designation refers to a reaction product between the ketone or aldehyde attached to any of the positions 1 to 22 with well-known ketone reagents, (carbazide, semi-carbazide, hydroxylamine, etc.)

Epoxy

This designation refers to an epoxy group attached to two nuclear carbon atoms.

Example of coding:
(a 3,4-epoxy pregnane)

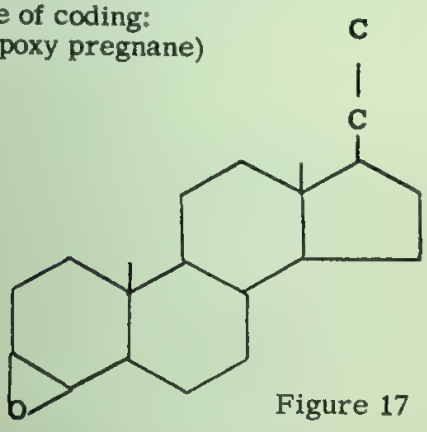


Figure 17

Codes: Column 35 Row 3 (epoxy) } Locations 3-4
Column 35 Row 4 (epoxy) }
Column 61 Row 0 (O-hetero compound)
Column 67 Row 4 (pregnane)

-O hydrocarbon

This designation refers to ethers or to any hydrocarbon attached to the nucleus through an oxygen atom.

Example of coding:
(a 3-methoxy pregnane)

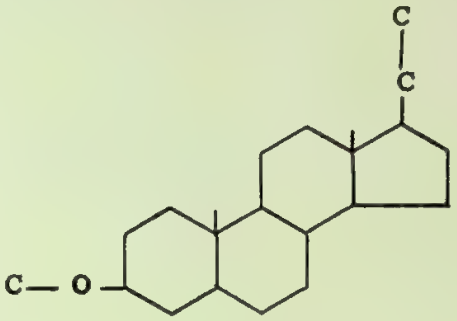


Figure 18

Codes: Column 37 Row 3 (-O hydrocarbon, location 3)
Column 67 Row 4 (pregnane)

-O acyl

This designation refers to an ester group attached to the steroid through the -O- atom of the group. Note that -O acyl is further defined by the 2B terms in column 60, rows 0 to 12. (See Figure 3).

Example of coding:
(a 21-acetyl pregnane)

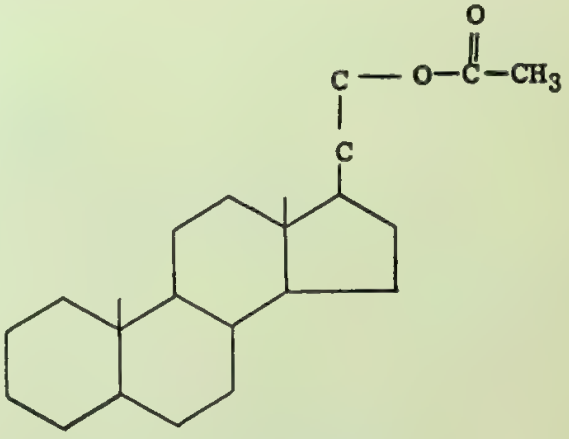


Figure 19

Codes: Column 40 Row 12 (-O acyl, location 21)
Column 60 Rows 0, 1, 5, 7 (O acyl and subgeneric terms)
Column 67 Row 4 (pregnane)

-O-Hetero

This designation refers to an oxygen-containing heterocyclic group. The heterocycle can be at-

tached through any of the atoms of the heterocycle at any one of the 22 positional locations.

Example of coding:
(a 3-furyl pregnane)

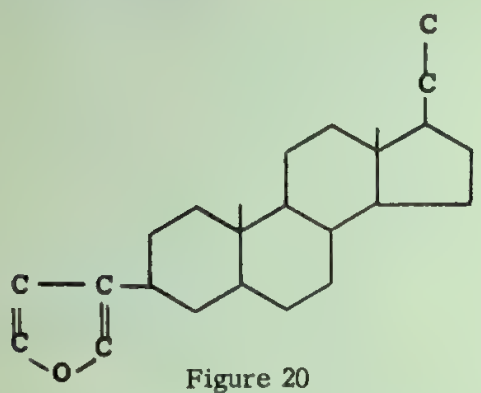


Figure 20

Codes: Column 41 Row 3 (O-hetero, location 3)
Column 61 Row 0 (O-hetero compound)
Column 61 Row 2 (O-hetero, furan)
Column 67 Row 4 (pregnane)

N-hetero

This designation refers to a nitrogen-containing heterocyclic group attached to any one of the 1 to 22 positions through any of the atoms of the heterocyclic group. Codes from 2B terms are also applied as in the following example.

Example of coding:
(a 3-pyrrole pregnane)

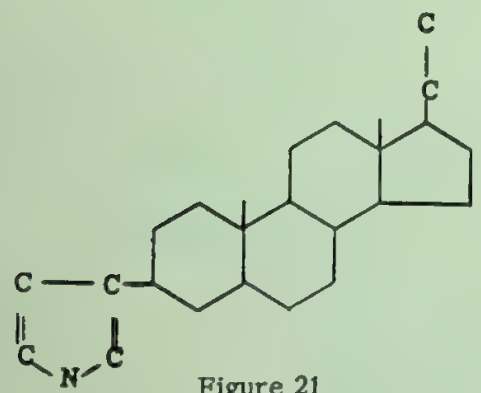


Figure 21

Codes: Column 43 Row 3 (-N-hetero, location 3)
Column 62 Row 0 (N-hetero compound)
Column 62 Row 5 (N-hetero, pyrrole)
Column 67 Row 4 (pregnane)

S-hetero

This designation refers to any sulfur-containing heterocyclic group attached to any of the 1 to 22 positions of the steroid molecule through any of the atoms of the heterocycle. Appropriate 2B terms are also applied.

Example of coding:
(a 20-thiophene pregnane)

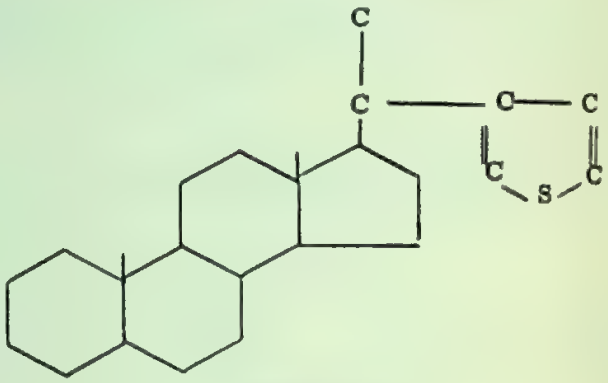


Figure 22

Codes: Column 46 Row 11 (-S-hetero, location 20)
Column 63 Row 0 (S-hetero compound)
Column 63 Row 1 (S-hetero, thiophene)
Column 67 Row 4 (pregnane)

Miscellaneous

This designation is the catch-all for any groups which are attached to any of the 1 to 22 positions and are not provided for above.

Definitions—2B Terms

The terms in 2B are, with few exceptions, not related to positions of substitution on the steroid nucleus. Several of the terms overlap 2A terms in concept. However, these overlap terms provide additional description of the group represented by a 2A term.

The 2B terms incorporate several levels of genericity. Indentations under any term provide further description or delineation of said term.

The terms are discussed below in the order of their allocated spaces on the IBM punched card.

Column 60

0. *O-Acyl.*—An ester group attached to the steroid nucleus through the -O- atom. The descriptors indented under O-Acyl provide greater specificity as to the *acid radical* of the ester group.

1. *Carboxylic.*—The carboxylic acid radical.

2. *Poly.*—The polycarboxylic acid radical.

3. *Unsat.*—An unsaturated acid radical. Aromatic unsaturation is excluded. Thus, the benzoic acid group is not coded as an unsaturated acid radical.

4. *Aromatic.*—The carboxylic acid radical containing an aromatic hydrocarbon group.

5. *Aliphatic.*—The aliphatic carboxylic acid radical.

6. *Subst.*—An aliphatic acid radical substituted by a non-hydrocarbon group.

7. *St. chain.*—A straight chain aliphatic carboxylic acid group with no substituents.

8. *Cycloalkyl.*—An aliphatic carboxylic acid group containing a cycloalkyl group.

9. *Branched.*—An aliphatic carboxylic acid group containing a branched alkyl group.

11. *Heterocyclic.*—A heterocyclic acid radical.

12. *Inorganic—except hal.*—Noncarboxylic inorganic acid radicals except halogen acid radicals.

Note.—By the multiple coding principle (*supra*, p. 5), all applicable terms are applied to a particular compound.

Example of coding:
(a 3-cyclohexyl formyl pregnane)

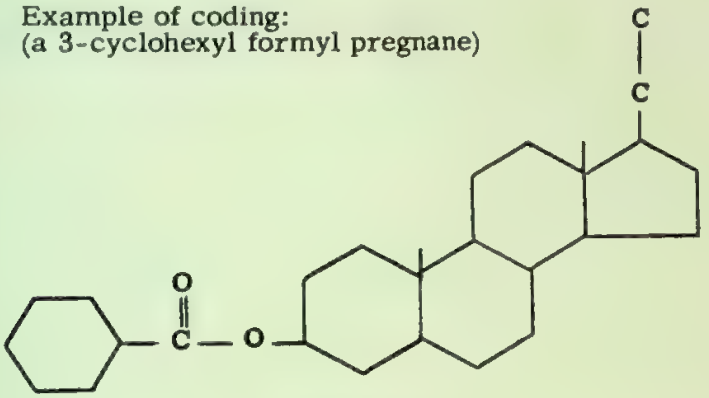


Figure 23

Codes: Column 39 Row 3 (-O acyl, location 3)
Column 60 Row 0 (O-acyl, compound)
Column 60 Row 1 (O-acyl, carboxylic)
Column 60 Row 5 (O-acyl, aliphatic)
Column 60 Row 8 (O-acyl, cycloalkyl)
Column 67 Row 4 (pregnane)
Column 67 Row 5 (21-unsubstituted)

Column 61

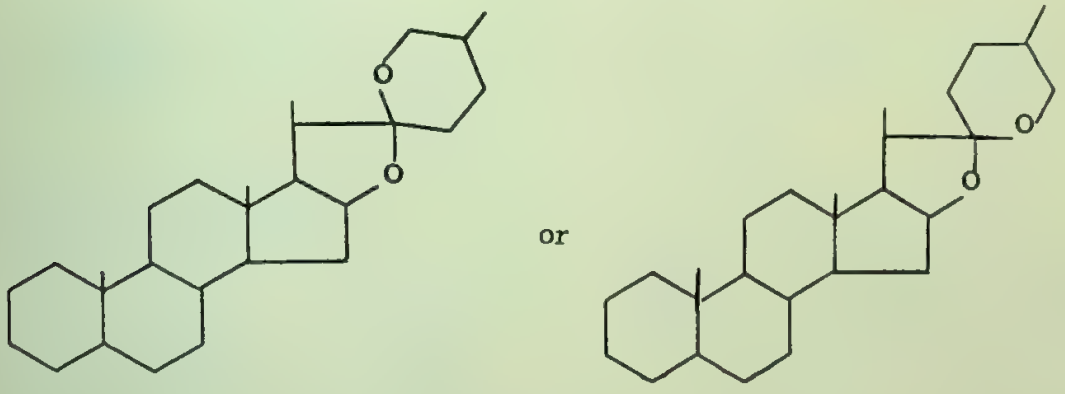
0. *O-hetero.*—An oxygen-containing heterocyclic group.

1. *Morpholine.*—The O-heterocycle in morpholine configuration.

2. *Furan.*—The O-heterocycle in furan configuration inclusive of saturation and unsaturation.

3. *Lactone.*—The O-heterocycle in lactone arrangment.

4. *Spirostane.*—The spirostane group normally found in sapogenins shown in Figure 24.



I II

Figure 24

5. ***Sub in O-spiro ring.***—Spirostanes containing substituents in the oxygen-containing ring other than hydrocarbon substituents.

6. ***Pseudosapo.***—The pseudosapogenins are acetic anhydride or other anhydride derivatives of spirostanes or sapogenins. A pseudosapogenin is shown in Figure 25.

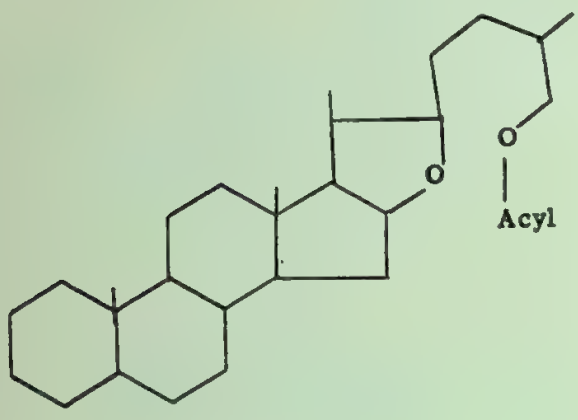


Figure 25

9. ***Misc.***—O-hetero steroids not specifically provided for by the terms above.

Column 62

0. ***N-hetero.***—Nitrogen-containing heterocyclic group.

1. ***Morpholine.***—The morpholine group. (The morpholine group is coded both as an O-hetero and an N-hetero group).

2. ***Piperidine.***—The piperidine nucleus.

3. ***Pyridine.***—The pyridine group which includes the dihydro and tetrahydro forms.

4. ***Pyrimidine.***—The pyrimidine group which includes the dihydro and tetrahydro forms.

5. ***Pyrrole.***—The pyrrole group which includes the saturated and unsaturated forms.

6. ***Thiazole.***—The thiazole group which includes the saturated and unsaturated forms.

9. ***Misc.***—Any nitrogen-containing heterocyclic groups not specifically provided for by the terms above.

Column 63

0. ***S-hetero.***—Sulfur-containing heterocyclic groups.

1. ***Thiophene.***—The thiophene ring which includes the saturated and unsaturated forms.

2. ***Thiazole.***—The thiazole group. (The thiazole group is coded both as an N-hetero and an S-hetero group).

9. ***Misc.***—Any sulfur-containing heterocyclic group not specifically provided for by the terms above.

Column 64

0. ***Bile cpds.***—This term includes bile acids, esters, and amides. The compounds are generally recognized by the presence of a chain of 3 to 5 carbons in the 17 position.

1. ***Acids.***—This term includes the bile acids, salts, amides, and esters.

2. ***Cholanic.***—The cholanic group.

3. ***Norcholanic.***—The norcholanic group.

4. ***Bisnorcholanic.***—The bisnorcholanic group.

Example of coding:
(a cholanic acid)

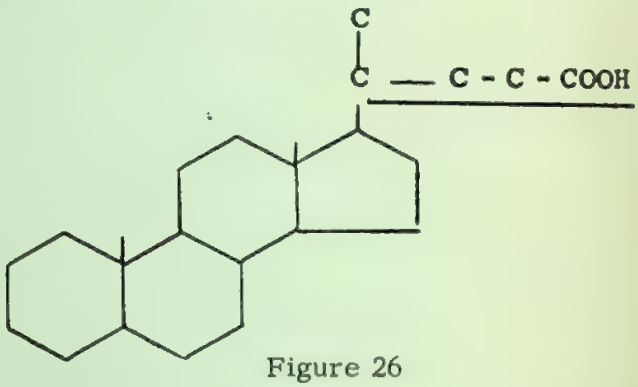


Figure 26

The underlined portion is coded as follows:

Codes: Column 64 Row 0 (bile compounds)
Column 64 Row 1 (bile compounds, acids)
Column 64 Row 2 (bile compounds, cholanic)

Column 65

0. ***Sterols.***—Sterols containing a hydrocarbon of more than 5 carbons in 17 position. The 2A hydrocarbon term does not apply to the sterol hydrocarbon side chain.

1. ***Ergosterol.***—The ergosterol nucleus.

2. ***Cholesterol.***—The cholesterol nucleus.

3. ***Vitamin D*** $_3$.—Vitamin D_3 has been included although it is not generally considered to be a steroid.

Column 66

0. ***Hal.***—A member of the halogen group.

1. ***Fl.***—Fluorine.

2. ***Br.***—Bromine.

3. ***I.***—Iodine

4. ***Cl.***—Chlorine

5-11. ***Double bonds.***—These terms provide a further delineation of the 2A double bond term, both broadly and specifically.

Column 67

1. ***Androstane.***—See *supra*, page 8.

2. ***Addition Compound.***—This term designates addition compounds such as amine salts, bisulfate addition products, etc.

3. *Maleic adducts.*—Steroid reaction products of maleic acid anhydrides or esters.

4. *Pregnane.*—See *supra*, page 8.

5. *21 Unsubstituted.*—Pregnanes which are not substituted in positional location 21.

6. *21 diazo.*—Steroids containing a diazo group in positional location 21.

9. *Misc.*—Any general term not specifically provided for.

CODING PROCEDURE

Instructions for Coding

1. Read the document for comprehension of the subject matter.

2. Encircle all of the codes representing both the 2A and 2B terms found in the patent on the coding form (Figures 2 and 3) in accordance with the principles of multiple coding and composite coding.

3. Extract all pertinent terms disclosed in the patent. The title, text, and claims of the patent are all parts of the disclosure for extraction and coding purposes. Chemical configurations disclosed as possible substituents as well as those more specifically disclosed are extracted and coded.

4. Have the coded information verified by a second individual. Note that the coded form represents a composite of all the substituents disclosed for the steroid nucleus in the given patent.

Example of Coding

The compound shown in Figure 27 is coded below. Applicable 2A and 2B terms are encircled on the coding forms presented in Figures 28 and 29, respectively. See pages 16 and 17.

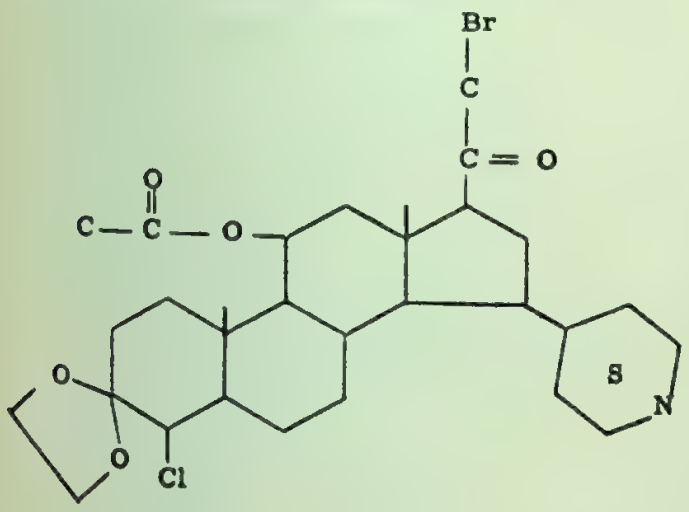


Figure 27

ALTERNATIVE CODING PROCEDURE—NON-COMPOSITE METHOD

A modification of the coding system described above can be employed when it is desired to provide machine selection and discrimination on an individual compound basis. This modification is in contrast with the method of composite coding (*supra*, page 5).

In individual compound coding, the 2A substituents and 2B terms disclosed for the particular compound are coded in the usual manner. The absence of 2A terms in the remaining positional locations on the steroid nucleus is indicated by employing the "H" descriptor for each of such locations. The only exception is in the 17 position, in which the punching of an "H" signifies the presence of only one substituent instead of two. When the keto group is in positional location 17, no "H" is punched. This device is not used for positional locations 20, 21 and 22.

To find those compounds which do not have double bonds in certain positions, the absence of a double bond is asked for by an appropriate wiring modification.

An illustration of individual compound coding is shown below for the compound progesterone which has the formula given in Figure 30.

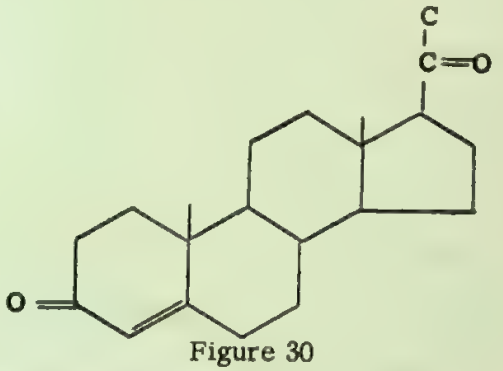


Figure 30

Codes: Column 1 Row 4
Column 15 Rows 1, 2, 4, 5, 6, 7, 8, 9
Column 16 Rows 1, 2, 4, 5, 6, 7, 8, 9
Column 21 Row 3,
Column 22 Row 11
Column 67 Rows 4, 5

PUNCHED-CARD FORMAT

The punched-card format is designed to accommodate steroid disclosures in both patents and published literature, domestic and foreign. The format is the same in all cards for columns 1 through 70 as follows (*supra*, page 5):

a. Columns 1-48 : 2A terms
b. Columns 60-67 : 2B terms
c. Columns 49-59, 68, 69: blank

For patent disclosures, the remaining ten columns are allocated as follows:

a. Column 70 : code signifying the Class 260 subclass in which the patent is classified; see the Appendix, Table I for particular codes.

Patent No. ..... Examiner ..... Punched .....

	0	1 2 3 4 5 6 7 8 9	10 11 12 13 14 15 16 17 18 19 20 21
═	22	1	2
∝ or allo	22	3	4
-C≡C	22	5	6
CH_3	22	7	8
CN	22	9	10
COOH or COOR	22	11	12
-C-sub	22	13	14
-H	22	15	16
NH_2 or N<	22	17	18
O H	22	19	20
=O	22	21	22
-Se-	22	23	24
-S-R	22	25	26
Hal	22	27	28
Hydrocarbon	22	29	30
Ketal	22	31	32
Ketone reagent	22	33	34
Epoxy	22	35	36
-O hydrocarbon	22	37	38
-O Acyl	22	39	40
-O-hetero	22	41	42
-N-hetero	22	43	44
S-hetero	22	45	46
Miscellaneous	22	47	48

Figure 28

60		61		62		63	
(0)	O-Acyl	(0)	O-Hetero	(0)	N-Hetero	0	S-Hetero
(1)	Carboxylic	1	Morpholine	1	Morpholine	1	Thiophene
2	Poly	2	Furan	(2)	Piperidine	2	Thiazole
3	Unsat.	3	Lactone	3	Pyridine	3	
4	Aromatic	4	Spirostane	4	Pyrimidine	4	
(5)	Aliphatic	5	Sub in O spiro ring	5	Pyrrole	5	
6	Substi.	6	Psuedosapo.	6	Thiazole	6	
(7)	St. chain	7		7		7	
8	Cycloalkyl	8		8		8	
9	Branched	9	Misc.	9	Misc.	9	Misc.
11	Heterocyclic	11		11		11	
12	Inorganic except hal						

64		65		66		67	
0	Bile cpds.	0	Sterols	(0)	Hal .	0	
1	Acids	1	Ergosterol	1	Fl.	1	Androstane
2	Cholanic	2	Cholesterol	(2)	Br.	2	Add. compd..
3	Norcholanic	3	Vitamin D_3	3	I	3	Maleic adduct
4	Bisnorcholanic	4		(4)	Cl. Double bond	(4)	Pregnane
5		5		5	5 (6)	5	21 Unsubstituted
6		6		6	5 (10)	6	21 Diazo
7		7		7	8 (9)	7	
8		8		8	8 (14)	8	
9	Misc.	9	Misc.	9	1 (2)	9	Misc.
11		11		11	1 (10)	11	
12		12		12		12	

Figure 29

b. Column 72 : code designating the country of patent origin; see the Appendix, Table II for particular codes.

c. Columns 74-80 : patent number.

d. Columns 71, 73 : blank

For published literature, the remaining ten columns are allocated as follows:

a. Column 71 : code representing the journal.

b. Column 72 : code designating the country of journal origin; see the Appendix, Table III for the combined journal–country codes.

c. Column 73 : year of journal; see the Appendix, Table IV for particular codes.

d. Column 74 : month or volume of journal, the month taking precedence. The volume is an arbitrarily assigned number—1 for 1958, 2 for 1959, etc.

e. Columns 75-78 : page numbers of journal article.

f. Columns 79, 80 : unique arbitrarily assigned number to each compound coded in the journal article.

USE IN PATENT SEARCHING

Use of the above-described system by Mechanized Examining Division A in examining steroid patent applications is outside the scope of this manual. Briefly, however, search questions are formulated based upon the patent applications in a manner similar to the method of coding the patents. Distinctions arise in the formulations based upon various factors including knowledge of the art and examination practice and procedures.

REFERENCES

1. Frome, Julius and Jacob Leibowitz. *A Punched Card System for Searching Steroid Compounds.* Patent Office Research and Development Report No. 7. Washington 25, D. C., Department of Commerce, 1956.

2. *Classification Bulletin, Class 260, Chemistry, Carbon Compounds.* No. 200, Rev. 1. U. S. Patent Office, Department of Commerce, Washington 25, D. C., 1956.

3. *Ibid.*, pp. 260-37, 260-38, 260-68 through 260-70.

4. Rosa, Manuel C. "Problems of Classifying Chemical Patents." *Journal of the Patent Office Society,* XXIX(1947), 241-261; *The Classification of Patents.* Rev. 2. U. S. Patent Office, Department of Commerce, Washington 25, D. C., 1946

Appendix

Table I

CODES DESIGNATING SUBCLASS ORIGIN

(Punch in Column 70)

Subclass	Row*
239.5	1
239.55	2
239.57	3
397	4
397.1	5
397.2	6
397.25	7
397.3	8
397.4	9
397.45	0
397.47	11
397.5	12

*Code is the same as the number of a particular row in column 70 since twelve subclasses are the subject of the puched-card deck.

Table II

CODES DESIGNATING NATIONAL ORIGIN

(Punch in Column 72)*

Code	Nation
S	Denmark
F	France
G	Germany
B	Great Britain
I	Italy
J	Japan
R	Russia and the USSR
H	Switzerland
U	United States

*No notation has been made in column 72 of the present punch-card deck to identify United States patents.

Table III

CODES IDENTIFYING JOURNAL AND NATIONAL ORIGIN

(Punch in Columns 71 and 72)

CODE		Journal name, origin
Col. 71	Col. 72	
C	D	Acta Chemica Scandinavica (Denmark)
A	D	Acta Endocrinologica (Denmark)
P	U	American Journal of Physiology (United States)
C	G	Angewandte Chemie (Germany)
E	F	Annales d'endocrinologie (France)
L	G	Annales der Chemie, Justus Liebigs (Germany)
B	B	Biochemical Journal (Great Britain)
M	B	British Medical Bulletin (Great Britain)
C	F	Bulletin de la societe chimique de France (France)
B	G	Chemische Berichte (Germany)
I	B	Chemistry and Industry (Great Britain)
B	F	Comptes rendus des seances de la societe de biologie et de ses filiales (France)
P	B	Current Chemical Papers (Great Britain)
D	R	Doklady Akademii Nauk Soyuza Sovetskikh Sotsialisticheskikh Respublik (Russia)
E	H	Experientia (Switzerland)
C	I	Gazzetta chimica italiana (Italy)
C	H	Helvetica Chimica Acta (Switzerland)
I	R	Izvestiya Akademii Nauk Soyuza Sovetskikh Sotsialisticheskikh Respublik. Otdelenie Khimicheskikh Nauk (Classe des sciences chimiques) (Russia)

Table III.—CODES IDENTIFYING JOURNAL AND NATIONAL ORIGIN—Con.

(Punch in Columns 71 and 72)

CODE		Journal name, origin
Col. 71	Col. 72	
B	J	Journal of Biochemistry (Japan)
B	U	Journal of Biological Chemistry (United States)
O	U	Journal of Organic Chemistry (United States)
C	U	Journal of the American Chemical Society (United States)
C	B	Journal of the Chemical Society (Great Britain)
C	J	Journal of the Chemical Society of Japan (Japan)
E	U	Journal of Clinical Endocrinology and Metabolism (United States)
N	B	Nature (Great Britain)
N	G	Naturwissenschaften (Germany)
P	J	Pharmaceutical Bulletin (Japan)
M	U	Proceedings of the Society for Experimental Biology and Medicine (United States)
S	U	Science (United States)
U	R	Ukrainskii Khimicheskii Zhurnal (Russia)
M	R	Voprosy Med. Khimii (Russia)
P	G	Zeitschrift fur physiologische Chemie (Hoppe-Seyler's) (Germany)
O	R	Zhurnal Obshchei Khimii (Russia)

Table IV

CODES DESIGNATING YEAR OF JOURNAL

(Punch in Column 73)

Year	Code
1958	11
1959	12
1960	0
1961	1
1962	2
1963	3
1964	4
1965	5
1966	6
1967	7
1968	8
1969	9

43386--U.S.Dept.of Comm--DC--1958